



AFRL-RY-WP-TR-2013-0218

OPEN-UNIVERSE THEORY FOR BAYESIAN INFERENCE, DECISION, AND SENSING (OUTBIDS)

**Stuart Russell and Shankar Sastry
University of California, Berkeley**

**Rene Vidal, I-Jeng Wang, Andreas Terzis, and Yanif Ahmad
The Johns Hopkins University**

**Avi Pfeffer
Charles River Analytics**

**Edwin Chong
Colorado State University**

**JANUARY 2014
Final Report**

Approved for public release; distribution unlimited.

See additional restrictions described on inside pages

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC)
(<http://www.dtic.mil>).

AFRL-RY-WP-TR-2013-0218 HAS BEEN REVIEWED AND IS APPROVED FOR
PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//

ANDREW FREEMAN, Program Manager
Electro-Optic Exploitation

//SIGNED//

CLARE MIKULA, Branch Chief
Electro-Optic Exploitation

//SIGNED//

DOUG HAGER, Deputy Chief
Layered Sensing Exploitation Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show “//Signature//” stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>						
1. REPORT DATE (DD-MM-YY) January 2014		2. REPORT TYPE Final		3. DATES COVERED (From - To) 29 September 2011 – 31 March 2013		
4. TITLE AND SUBTITLE OPEN-UNIVERSE THEORY FOR BAYESIAN INFERENCE, DECISION, AND SENSING (OUTBIDS)				5a. CONTRACT NUMBER FA8650-11-1-7153		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER 61101E		
6. AUTHOR(S) Stuart Russell and Shankar Sastry (University of California, Berkeley) Rene Vidal, I-Jeng Wang, Andreas Terzis, and Yanif Ahmad (The Johns Hopkins University) Avi Pfeffer (Charles River Analytics) Edwin Chong (Colorado State University)				5d. PROJECT NUMBER 1000		
				5e. TASK NUMBER 11		
				5f. WORK UNIT NUMBER YOPE		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California, Berkeley Berkeley, CA 94720				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command United States Air Force				The Johns Hopkins University, Baltimore, MD 21287 Charles River Analytics, 625 Mt Auburn St., Cambridge, MA 02138 Colorado State University		
				10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/Ryat		
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RY-WP-TR-2013-0218		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.						
13. SUPPLEMENTARY NOTES This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.						
14. ABSTRACT This report describes work in Phase I of the OUTBIDS project under the DARPA MSEE program. The goal of OUTBIDS was to develop the theoretical and technological foundations for sensor data interpretation as a form of probabilistic inference. Achieving this goal requires a representation formalism for probability models of sufficient expressive power to handle the complexity of real-world sensor data. The problem involves two primary sources of difficulty: first, the underlying world generating the data typically contains many initially unknown objects interacting over time in complex ways; second, the mapping from objects and behaviors to sensor data is itself (as in the case of visual perception, for example) very complex. The core of the project is the BLOG (Bayesian LOGic) language, which combines probabilistic semantics with the expressive power of first-order logic. Unlike other attempts to combine probability and logic, BLOG supports open-universe models, which allow for uncertainty over the existence and identity of objects. We believe this is a prerequisite for any probabilistic approach to general perception. The team made substantial progress on developing and refining the BLOG language by writing a broad range of models, including two models for computer vision tasks (adaptive video background subtraction for object tracking and a simple form of 3D object recognition and scene reconstruction). We also made good progress towards an efficient inference engine, including new algorithms and initial work on compiler and parallelization technology for BLOG inference. We showed that BLOG could be extended to open-universe decision models and developed algorithms for sensor planning on this basis. We also developed a new theoretical framework for utility-directed inference and proved several foundational theorems. In addition to developing BLOG, we also developed a range of generative probability models specifically for visual perception, in both 2D and 3D, as well as a new family of inference algorithms for these models. We demonstrated superior performance on several benchmark data sets. The idea was that these models would be developed and tested in parallel to the BLOG substrate and gradually migrated to BLOG as the technology matured. As noted above, we developed one 3D vision system entirely within BLOG as a proof of concept. We continue to believe that the overall approach can succeed in providing a new foundation and technological capability for machine perception. Research on BLOG will continue under the DARPA PPAML program. Meanwhile, the United Nations has submitted for approval by the member states a BLOG-based global seismic monitoring system for the Comprehensive Nuclear-Test-Ban Treaty.						
15. SUBJECT TERMS information extraction, incremental learning, classification, decision making, computer vision, knowledge representation						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 22	19a. NAME OF RESPONSIBLE PERSON (Monitor) Andrew M. Freeman	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include Area Code) N/A	

Table of Contents

1	Executive Summary	1
2	Task T1 - Mathematical formalism	2
3	Task T2: Inference algorithms	4
4	Task T3 - Sensor planning.....	6
5	Task T4 - Sensor processing.....	7
5.1	Probabilistic Models of 3D Objects.....	7
5.2	Probabilistic Models of 2D Objects.....	9
5.3	Single-Sample Face Recognition via Sparse Illumination Transfer.....	11
5.4	Compressive Shift Retrieval.....	12
5.5	Implementation of Vision Models in BLOG.....	12
6	Tasks T5 & T6 – Software implementation & Experimental evaluation	14
7	References	15
	List of Acronyms, Abbreviations, and Symbols	17

1 Executive Summary

This report describes work in Phase I of the “Open-Universe Theory for Bayesian Inference, Decision, and Sensing” (OUTBIDS) project under the Defense Advanced Research Projects Agency (DARPA) Mathematics of Sensing, Exploitation, and Execution (MSEE) program. The goal of OUTBIDS was to develop the theoretical and technological foundations for sensor data interpretation as a form of probabilistic inference. Achieving this goal requires a representation formalism for probability models of sufficient expressive power to handle the complexity of real-world sensor data. The problem involves two primary sources of difficulty: first, the underlying world generating the data typically contains many initially unknown objects interacting over time in complex ways; second, the mapping from objects and behaviors to sensor data is itself (as in the case of visual perception, for example) very complex.

The core of the project is the Bayesian LOGic (BLOG) language, which combines probabilistic semantics with the expressive power of first-order logic. Unlike other attempts to combine probability and logic, BLOG supports *open-universe* models, which allow for uncertainty over the existence and identity of objects. We believe this is a prerequisite for any probabilistic approach to general perception.

The team made substantial progress on developing and refining the BLOG language by writing a broad range of models, including two models for computer vision tasks (adaptive video background subtraction for object tracking and a simple form of 3D object recognition and scene reconstruction). We also made good progress towards an efficient inference engine, including new algorithms and initial work on compiler and parallelization technology for BLOG inference. We showed that BLOG could be extended to open-universe decision models and developed algorithms for sensor planning on this basis. We also developed a new theoretical framework for utility-directed inference and proved several foundational theorems.

In addition to developing BLOG, we also developed a range of generative probability models specifically for visual perception, in both 2D and 3D, as well as a new family of inference algorithms for these models. We demonstrated superior performance on several benchmark data sets. The idea was that these models would be developed and tested in parallel to the BLOG substrate and gradually migrated to BLOG as the technology matured. As noted above, we developed one 3D vision system entirely within BLOG as a proof of concept. We continue to believe that the overall approach can succeed in providing a new foundation and technological capability for machine perception. Research on BLOG will continue under the DARPA PPAML program. Meanwhile, the United Nations has submitted for approval by the member states a BLOG-based global seismic monitoring system for the Comprehensive Nuclear-Test-Ban Treaty.

2 Task T1 - Mathematical formalism

In this project, the primary effort in T1 was focused on developing the mathematical formalism for the BLOG language and ensuring its adequacy for a wide range of tasks including perception tasks.

- PhD student Nick Hay, postdoc Siddharth Srivastava, and PI Russell developed a generalized measure-theoretic representation theorem (unpublished) for probability models defined as products of conditional models, which will form the basis for the semantic foundation of BLOG with real-valued functions and functions of real arguments.
- Postdoc Lei Li, PI Russell, and several undergraduate students developed roughly 20 new models, including
 - Tug of war (a test case requested by DARPA PPAML program manager Kathleen Fisher.
 - Infinite-state hidden Markov model
 - Weather forecasting
 - Birthday “paradox”
 - Multitarget tracking with detection failure, false alarms, track initiation and termination
 - Population Estimation (Urn-Ball)
 - Sybil Attacks
 - A simple probabilistic context-free grammar for natural language
 - Document-topic model (latent Dirichlet allocation)
 - Citation information extraction
 - Students/courses/grades (with first-order quantifiers)
 - Adaptive video background subtraction
- PI Russell and postdoc Srivastava extended BLOG to allow for perception and action and the definition of first-order open-universe MDPs and POMDPs (see Task T3). Two such models were
 - Partially observed Monopoly
 - One player Blackjack

CRA co-PI Pfeffer similarly extended the probabilistic programming (PP) approach to allow for decision making [23]. In both BLOG and PP approaches to decision making, the information on which the decision is based can be a complex data structure, such as a social network or DNA sequence, and the process of generating outcomes given decisions can also be complex. To plan for a decision that can potentially be applied in a very large or infinite number of situations, we developed a sampling algorithm combined with a nearest neighbor approach to determine strategies.

- In the process of developing these models we improved the syntax and semantics of BLOG to better match those models (along with modifying the inference engine to handle the revised language). The updated syntax is explained in a new version of the BLOG manual [1]. New developments include nested if-then-else statements; a uniform

interface for both fixed and random arguments of functions; map data structures; general algebraic expressions for defining random functions; and vector- and array-valued random variables (so that, for example, linear-Gaussian systems with unknown covariances can be modeled).

- The BLOG language and inference engine were made available at <http://bayesianlogic.cs.berkeley.edu>, as both a downloadable system and as an interactive web interface to the BLOG engine running at Berkeley.

PI Russell's work on open-universe probability models was recognized by two significant honors: the *Chaire Blaise Pascal*, France's highest award for foreign scientists in any discipline, and the senior *Chaire d'Excellence* of France's Agence Nationale de la Recherche. In addition, he presented the research program in a number of distinguished invited lectures throughout Europe. Postdoc Lei Li participated in the 2012 Young Investigator Conference at Stanford, and gave an invited talk about Open Universe Probabilistic Models and BLOG language. He also participated in an ISAT workshop on Probabilistic Programming, which led to the DARPA PPAML program (in which Berkeley is a funded participant).

Contributors

UC Berkeley: Stuart Russell, Lei Li, Aaron Wong, Akihiro Matsukawa, Chenglai Victor Huang, Dan Wang

Charles River Associates: Avi Pfeffer

3 Task T2: Inference algorithms

In the second task, we worked on efficient inference algorithms for open universe probabilistic generative models. Pre-existing algorithms included likelihood weighting and a Metropolis-Hastings MCMC algorithm based on “parental sampling” (in which variables are sampled, conditioned only on their parents’ values). These algorithms while provably complete and correct, are often slow to converge. Models with deterministic or near-deterministic relationships and temporal models with static variables are particularly problematic (a fact noted by many other developers of probabilistic inference systems). We have developed several fast algorithms for these challenging inference problems.

The first algorithm is for sampling a block of variables with deterministic relationships specified in BLOG. With these deterministic relationships, standard sampling algorithms such as single-variable Gibbs sampling and Metropolis-Hastings with parental sampling fail completely. We developed solutions initially for a canonical case, where one variable is constrained to be the sum of k other variables. For the case where each random variable has a bounded range of integer values, we derived an $O(k)$ dynamic programming algorithm called *exact constrained sequential sampling* (ECSS). For the more general, continuous case, we proved NP-hardness of the sampling problem and proposed a *dynamic scaling* algorithm (DYSC), a form of sequential importance sampling, which works well in practice. The schemes of ECSS and DYSC can be easily generalized to constraints expressible as arbitrary trees of invertible binary [2].

For models of temporal processes, the biggest open problem in the field is that of devising practical inference algorithms for models with static (atemporal) variables. This case includes models with unknown parameters, which become static variables in the Bayesian formulation. Standard “sequential Monte Carlo” inference algorithms such as particle filtering fail for all such models, which violate the conditions required for the algorithm’s convergence. We addressed the problem first in the context of dynamic Bayesian networks (DBN), as a prelude to tackling the same issue in DBLOG (the temporal extension of BLOG). We developed a more general form of Storvik’s filter, which uses fixed-dimensional sufficient statistics to maintain the Gibbs distribution for the static variables. We generalized the family of distributions to which the algorithm can be applied and showed that a Taylor approximation to any analytic Gibbs distribution satisfies these conditions; the following paper on the new algorithm, the *extended parameter filter* [3], was accepted at ICML 2013. Our new algorithm takes only constant running time per update, as oppose to linear time in traditional SMC.

In the process of considering temporal probabilistic models for complex processes, we developed two new families of models and associated algorithms. The first allows for *uncertain observation times*, which are common in real data – for example, in clinical observations entered after the fact; we showed that such data can be handled using a novel dynamic programming algorithm [6]. The second allows for tensor data, in which observations at a given time step exhibit multidimensional (e.g., spatial) structure (as opposed to one-dimensional vector models such as Kalman filters).

We developed a dynamical tensor model that gives far better estimation and system-identification results than the standard vectorization approach, as demonstrated on both simulated data and on a real-world climate data set. A preliminary paper was published at the 2012 Workshop on Understanding Climate Change [4] and a full paper has been accepted at NIPS 2013. Undergraduate student Sharad Vikram, working with Lei Li, developed a temporal model and inference algorithm for 3D gesture recognition, published at CHI 2013 [5].

In the area of utility-directed inference, originally a principal goal of the MSEE program and OUTBIDS project, we formulated a general theoretical framework for metalevel decisions (i.e., decisions about computations); disproved the commonly held belief that bandit theory was the applicable framework; proved the first general theorems concerning non-negative expected utility and eventual termination¹ of optimal metalevel policies; and devised a new heuristic approximation for max and min nodes in lookahead trees (nested selection problems). A paper was accepted for oral presentation at UAI 2012 (8% accept rate) [7].

Contributors

UC Berkeley: Stuart Russell, Lei Li, Shaunak Chatterjee, Nick Hay, Mark Rogers, Yusuf Erol, Bharath Ramsundar

¹ We also proved the counterintuitive result that optimal computational policies *can* in some cases compute forever, incurring infinite cost for a decision of finite value.

4 Task T3 - Sensor planning

Sensor planning was originally a principal goal of the MSEE program. Our work in this area was carried out in the theoretical framework of partially observable Markov decision processes (POMDPs), wherein optimal sensing decisions can be made based on current and future *belief states* of the agent. The utility associated with a sensor modality or a specific sensing action can be predicted by estimating the performance deviation between solutions of the POMDPs with and without the sensor modality/action. We illustrated such a formulation and its use for utility prediction using a prototyped problem with notional cameras and an airborne sensor.

Existing frameworks for POMDPs do not adequately express open-universe POMDPs, or problems where the agent needs to use its sensors to infer the existence of entities, reason about their identities and properties, and select appropriate actions. While existence uncertainty is adequately captured in the BLOG system using a first-order probabilistic language, modeling an agent's sensor and actuator capabilities in any first-order system leads to significant problems. The “obvious” approach to modeling the capabilities of a sensor or effector leads to provably false claims and incorrect specifications of the decision process. To overcome this technical difficulty, we developed a mathematical formulation of the open-universe POMDP problem for use in BLOG [8]. This formulation utilizes ideas from modal logic to enable the expression of sensor and actuator capabilities consistent with their true physical properties. We used this framework to develop algorithms for evaluating the expected values of two types of sensor-planning policies: those expressed as finite-state machines (mapping patterns of observations to actions) and those expressed as belief-state-query policies (mapping first-order belief states, computed by DBLOG queries, to actions). Finally, we conducted policy evaluation for various finite-state machine policies in the target detection domain.

To establish the theoretical justification for policy rollout with belief-state-query policies, we investigated possible theoretical guarantees for asynchronous policy iteration with general feature-based policies. We derived a condition on the features that is sufficient to ensure policy improvement for feature-based policy rollout [9]. The implication of this preliminary theoretical result on the proper structure of belief-state-query policies will be part of our future research.

Contributors

UC Berkeley: Stuart Russell, Siddharth Srivastava

Johns Hopkins University: I-Jeng Wang, Anshu Saksena

Charles River Associates: Avi Pfeffer

Colorado State University: Edwin Chong

5 Task T4 - Sensor processing

5.1 Probabilistic Models of 3D Objects

We developed three different kinds of probabilistic models of 3D objects, together with appropriate inference and learning techniques, for solving a variety of scene interpretation tasks such as 3D object detection, 3D reconstruction, 3D pose estimation, object categorization, and 2D object segmentation from a single 2D image.

3D reconstruction, pose estimation and categorization using hypothesize and bound. In [15,16,17], we developed a probabilistic generative object model that incorporates prior knowledge about the 3D shapes of different object classes to solve the joint problems of 3D reconstruction, pose estimation and object categorization from a single 2D image. The generative process involves sampling a 3D shape from a volumetric prior model for each object class K . A geometric transformation T consisting of a 3D rotation, translation and scaling is then applied to the 3D shape. The transformed 3D shape is then projected onto the image to produce a 2D shape. The 2D shape is then used to generate the observations, which are represented by a 2D foreground probability map. The likelihood $L(H)$ of this generative model measures how well a hypothesis $H = (K, T)$ “explains” the input image.

Given a background-subtracted image, we developed an efficient algorithm for finding the hypotheses that maximize the likelihood. This problem is very challenging because of two reasons. First, the space of all possible hypotheses is huge. Second, the evaluation of the likelihood is very expensive, because it involves integrating over all the voxels in the 3D shape and the pixels in the 2D shape. To address these challenges, we developed an efficient inference algorithm called hypothesize-and-bound (H&B). H&B is a generic optimization algorithm designed precisely for cases where the number of hypothesis is very large and the cost of evaluating $L(H)$ for each hypothesis is very high. The core idea behind H&B is to discard suboptimal hypotheses with as few computations as possible. This is done by computing very cheap upper and lower bounds for $L(H)$. A hypothesis is discarded if its upper bound is smaller than the lower bound for another hypothesis. The active hypotheses (the ones that have not been discarded) are ranked (e.g., by the margin between the upper and lower bound) and the bounds for the best-ranked hypotheses are refined. The refinement process is based on a coarse-to-fine representation of 3D shapes that we had proposed in [14]. A finer shape leads to tighter bounds, which allow one to disregard more suboptimal hypothesis. The tightness of these bounds is hence a function of the number of cycles spent on refining the hypothesis H . The steps of bound refinement and hypothesis pruning are then repeated until the set of active hypothesis cannot be further pruned.

We performed classification and pose estimation experiments with object classes like plates, cups, mugs and bottles, first in a controlled setting where simple background subtraction is applicable. These experiments demonstrated the efficiency and accuracy of the proposed approach. However, a weakness of this approach is that it is not applicable to cluttered scenes,

because the probabilistic model uses a background-subtracted image as an input. One approach that we did not explore is to use more sophisticated appearance models and then apply H&B to the new likelihood. Instead, we kept the generative model intact and used powerful discriminative models for 2D object detection and interactive image segmentation to generate a background-subtracted image. Specifically, we used the 2D object detector of Felzenszwalb et al. to find candidate regions where the object may be present. The resulting detection scores were used to seed the interactive segmentation algorithm by Grady, which produces the background-subtracted image needed by the H&B algorithm. We performed classification and pose estimation experiments on images of plates and mugs placed on a dining table in a real setting that contains a large amount of clutter. Our experiments showed that the proposed approach achieved a classification performance of about 80-90% and pose estimation errors of about 3-5 cm.

3D detection and pose estimation using 3D deformable part models. We developed a probabilistic generative object model composed of deformable 3D parts with a joint spatial distribution specific to each category and consistent appearances in part-specific characteristic viewpoints to solve the joint problems of 3D object detection and pose estimation in a single 2D image. For a given object category, our model captures the generative process from the 3D scene content, up to observable 2D image features with appropriate priors on (i) the joint 3D configuration of deformable object parts; and (ii) the appearance of each part captured in their canonical pose, which we denote as *aspect*. More specifically, the proposed model captures the 3D pose of an object relative to the world coordinate system. The object is decomposed into a collection of planar parts, or aspects, which represent a canonical view of the object. A probabilistic model captures the arrangement of these parts relative to the object coordinate system in 3D. Specifically, the shape and location of each part is modeled with a Gaussian distribution, while its orientation is modeled with a Von Misses distribution. The appearance of each part is modeled in two different ways. In one approach, the appearance of each part is modeled using probabilistic PCA applied directly to the image intensities. In another approach, the appearance of each part is modeled using a Histogram of Oriented Gaussian (HOG) template whose entries follow a log-normal distribution. The appearance of the visible parts of the object is then projected onto the image. While the observation model is also based on HOG, the relationship between observed HOG entries, and the latent 3D geometry and appearance is nontrivial. One key contribution was to derive a linear relationship between observed HOG entries in 2D and latent HOG entries in 3D under the affine projection model. This leads to a novel generative model for relating 2D and 3D information that can cover a continuous range of poses, naturally handles self-occlusions among object parts, and links the latent world structure and texture to individual pixels on the observed image via camera projection. This model has many advantages with respect to the state of the art. First, unlike prior work that separates the modeling of 3D geometry and appearance, we proposed a unified model for both the geometry and appearance of an object class. Second, unlike prior work that requires annotations for object views or object parts, our work treats both the geometry and appearance of object parts as latent

variables of the model, which are estimated during inference. Third, unlike prior work that learns different pieces of the model independently, use matching between 3D models and 2D views and/or voting to determine the object pose, our model is generative, and hence amenable to probabilistic learning and inference.

We implemented the MCMC inference algorithm, which is designed to efficiently sample the 3D pose and appearance variables conditional to an observed image. Using hardwired parameters, we were able to run the MCMC inference on the PASCAL dataset for cars with good pose estimation results. However, inference was relatively slow and there is a need for developing more efficient methods. We also implemented an Expectation Maximization (EM) algorithm for learning the parameters of this model from training images. However, the inference algorithm was too slow to be able to learn the model parameters from a large dataset. Moreover, good initialization of both the model parameters and of the latent variables was critical for the success of the method. This motivated the development of discriminative methods that can be used for seeding the generative approach, as described next.

3D detection and pose estimation using wireframe models, and branch and bound. In [18], we developed an alternative approach to solving the same problems, in which 3D object templates of edge primitives are used as prior information. A 3D “wireframe” model is acquired effectively as a mean shape from 2D object blueprints, which contain orthographic and canonical views that can be easily registered to an accurate reconstruction without the need of solving correspondence matching. Given this model, the objective is to determine the continuous pose of the object such that the projection of the 3D model primitives best matches the observed edges of the 2D input image. The optimal set of hypotheses is found using a branch-and-bound algorithm applied to the pose space. The inference algorithm uses HOG features of the image as input, and for a given pose range, it efficiently computes tight upper bounds of the matching score as the sum of HOG entries at camera-projected locations and orientations of model primitives. For this, 3D integral images of quantized HOGs are employed in a novel way to evaluate in constant time the maximum attainable scores of individual primitives. In a priority queue determined by their score upper bounds, subsets of the pose space are sequentially processed and divided into finer cells until a finest resolution is reached at which point the optimum is achieved. We experimented with this method for localizing and estimating poses of sedan cars on two publicly available datasets, from Savarese *et al.*, and from PASCAL VOC 2006, and achieved results better than the state-of-the-art with testing times as low as less than half a second.

5.2 Probabilistic Models of 2D Objects.

We developed three different conditional random field models, together with appropriate inference and learning techniques, for solving a variety of scene interpretation tasks such as 2D object detection, object categorization, and 2D object segmentation from a single 2D image.

2D segmentation and categorization using conditional random field models. In [19] we developed a Conditional Random Field (CRF) model for joint categorization and segmentation

of objects in 2D images. Prior work addressed this problem using bottom-up approaches where local information extracted from a pixel (or superpixel) is used to define the unary potentials, which define the cost of assigning each pixel (or superpixel) to each one of the object categories. The pairwise potentials capture the cost of assigning two category labels to two pixels (superpixels) and are designed to encourage spatial coherence. While higher-order potentials that capture longer range interactions for label consistency had been proposed, such higher-order potentials are defined over pre-defined regions and are unable to capture long-range interactions over the whole region occupied by an object. Our work proposed a new class of higher order potentials for joint categorization and segmentation, which encode the cost of assigning a label to large regions in the image. Such potentials are defined as the output of a classifier applied to the histogram of all the features in an image that get assigned the same label. These histograms are effectively a Bag-of-Features (BoF) representation for a region of the image containing an object category. These top-down potentials obtained from the BoF representation can seamlessly be integrated with traditionally used bottom-up potentials, thus providing a natural unification of global and local interactions. The parameters for these potentials can be treated as parameters of the CRF and hence be jointly learnt along with other parameters of the CRF. For this purpose, we proposed a novel framework for learning classifiers that are useful for categorization as well as for multi-class segmentation. Experiments on the Graz dataset showed that our framework improves the performance of multi-class segmentation algorithms.

2D segmentation and categorization using latent conditional random field models. One disadvantage of the approach in [19] is that the parameters of the classifiers are learned independently from the visual words used by the BoF representation. To address this issue, in [20] we proposed a latent CRF model where the observed variables are category labels and the latent variables are visual word assignments. The CRF energy consists of a bottom-up segmentation cost, a bag of (latent) words categorization cost, and a dictionary learning cost. Together, these costs capture relationships between image features and visual words, relationships between visual words and object categories, and spatial relationships among visual words. The segmentation, categorization, and dictionary learning parameters are learned jointly using latent structural SVMs, and the segmentation and visual words assignments are inferred jointly using graph cuts. Evaluation on the Graz02 dataset showed improvements in two categories (bicycle and cars) and no improvement on people.

2D detection, segmentation and categorization using conditional random field models. In recent years, automatic human pose detection and segmentation in images and videos has become increasingly important with applications ranging from activity recognition in security camera networks to automatic sports analysis. Estimating the pose and location of humans (or other articulated objects) in static images is generally a hard problem as the background pixels around the human is unknown a priori, as is the configuration of his/her joints.

In our work [10], we have proposed a new formulation to bridge the gap between segmentation algorithms that consider local neighborhoods and categorization algorithms that consider non-

local neighborhoods. The method uses models of objects with deformable parts, classically used for object categorization, to solve the joint categorization and segmentation problem. We used these models to introduce two new classes of potential functions: The first class of potential functions encode the model score for detecting an object as a function of its visible parts only, and the second class encode shape priors for each visible part and is used to bias the segmentation of the pixels in the support region of the part, towards the foreground object label.

We have shown that most existing deformable parts formulations can be used to define these potential functions and that the resulting potential functions can be optimized exactly using min-cut. As a result, these new potential functions can be integrated with most existing random-field based formulations for joint categorization and segmentation.

Coarse-to-fine semantic video segmentation using supervoxel trees. Another disadvantage of the approach in [19] is that inference becomes extremely slow as the number of nodes (pixels or superpixels) increases. This is particularly the case when we wish to perform joint segmentation and categorization of videos. To address this issue, in [21], we proposed a coarse to fine approach for inference in hierarchical CRFs. Our strategy is based on a hierarchical abstraction of the supervoxel graph that allows us to minimize an energy defined at the finest level of the hierarchy by minimizing a series of simpler energies defined over coarser graphs. The strategy is exact, i.e., it produces the same solution as minimizing over the finest graph. It is general, i.e., it can be used to minimize any energy function (e.g., unary, pairwise, and higher-order terms) with any existing energy minimization algorithm (e.g., graph cuts and belief propagation). It also gives significant speedups in inference for several datasets with varying degrees of spatio-temporal coherence.

5.3 Single-Sample Face Recognition via Sparse Illumination Transfer

Face recognition has been a classical problem in pattern recognition literature. The community’s sustained interest in this problem is mainly due to two reasons. First, in face recognition, we encounter many of the common variabilities that plague image-based pattern recognition systems in general: illumination, occlusion, pose, and misalignment. Second, face recognition has a wide spectrum of practical applications. If we could construct an extremely reliable automatic face recognition system, it would have broad implications for identity verification, access control, security, and online image search.

In 2009, inspired by the emerging compressive sensing theory, we proposed a new face recognition framework called *sparse-representation based classification* (SRC), which can successfully address many of the image nuisances above. The framework is built on a subspace illumination model characterizing the distribution of a corruption-free face image under a fixed pose, one subspace per each subject class. Therefore, when an unknown query image is jointly represented by all the subspace models, only a small subset of these subspace coefficients need to be nonzero, which would primarily correspond to the subspace model of the true subject. Since the publication of this work, the SRC framework has been well received as a breakthrough in face recognition that deals with high-resolution, high-noise images.

In this project, we have extended the scope of the SRC framework to one of the most challenging scenarios known as single-sample face recognition [12], whereby the test image may still undergo unknown illumination change, pose variation, and pixel corruption, but the training image set only contains one training image per subject class. Furthermore, we consider the condition where the illumination of the single training images is also varying. To compensate the missing illumination information, a sparse illumination transfer technique was introduced. By enforcing a sparse representation of the query image, the method can recover and transfer the missing pose and illumination information from the alignment stage to the recognition stage using a set of additional illumination examples of irrelevant face images.

Our extensive experiments have demonstrated the new algorithms significantly outperform the existing algorithms in single-sample regime and with fewer restrictions. A benchmark of the performance is shown in Table 1 below. UC Berkeley has filed a US patent application [13].

Table 1. Single-sample face recognition comparison on the Multi-PIE database.

Method	Session 1 (%)	Session 2 (%)
DSRC (Wagner et al. 2012)	36.1	35.7
MRR (Yang et al. 2012)	46.2	34.6
SIT	79.9	65.7

5.4 Compressive Shift Retrieval

In signal processing, shift retrieval is a fundamental problem that concerns the recovery of a time shift that relates two signals. The technique has been used in many applications such as active sonar and target triangulation. Traditionally, the shift retrieval problem is solved by maximizing the cross-correlation between the two signals.

In our work [11], we have developed a compressive variant where the measurement of the signals is undersampled. While the standard procedure to shift retrieval in this case calls for the recovery of the signal itself, e.g., using compressive sensing techniques. We stipulate that the shift can be exactly recovered directly from the compressed measurements if the sensing matrix satisfies certain conditions. A special case is the partial Fourier matrix. In this setting, we show that the true shift can be found by as low as one measurement per signal. We further showed that the shift can also be recovered when the measurements are perturbed by noise.

5.5 Implementation of Vision Models in BLOG

We implemented a proof of concept version of and-or trees in BLOG. The structure was as follows. A given scene could be a street scene or a farm scene. A street scene contained bikes and people. A farm scene contained people and carts. Bikes and carts both were allowed to have

wheels, but each bike was constrained to have 2 wheels and each cart was constrained to have 3 wheels. Given this model, we provided evidence for various levels of the tree to see if the correct inference was made. For example, if we observed 4 wheels, then the BLOG inference suggested that with probability close to 1, the given scene was a street scene, which had two bikes. Similarly, if we observed 9 wheels, then the BLOG inference suggested that with probability close to 1, that the given scene was a farm scene with three carts. We also tested the case where the observed scene had 6 wheels (i.e., 3 bikes or 2 carts), and noted that the ambiguity of the type of scene was reflected in the inference results.

We also implemented a simple scheme for the projection of the 3D world to the 2D image. Specifically, we defined the extent of the world in 3D co-ordinates and chose a subset of the 3D voxels as being occupied, i.e., voxels where an object would be present. Given a 2D pixel in the image, we consider all the 3D voxels that lie on the back-projected ray that passes through the pixel and assign the color of the pixel as the color of the occupied voxel on this ray that is closest to the camera. If there is no occupied voxel on this ray, we assign some other color, which would be a property of the background. For example, in an outdoor scene, one would see the color of the sky at pixels whose corresponding rays do not pass through any occupied voxel. Finally, we also considered the problem of generating objects in the 3D world such that they do not overlap. To consider a simplistic scenario, we considered the problem of generating several non-intersecting intervals on the real line. In order to do this, we noted that we could define a predicate that depends on the boundaries of the intervals and is false when the intervals intersect. By providing evidence that this predicate is true, we ensure that a valid world corresponds to one where the intervals don't intersect.

Contributors

UC Berkeley: Sankar Sastry, Allen Yang, Dheeraj Singaraju, Nikhil Naikal

Johns Hopkins University: Rene Vidal, Erdem Yoruk, Aastha Jain, Siddharth Mahendran

6 Tasks T5 & T6 – Software implementation & Experimental evaluation

Our efforts in the first phase across UCB and JHU focused on API design and implementation on the Java codebase for BLOG. We worked on four areas, namely:

- i. The decoupling of sampling algorithms from the underlying graph data structures representing BLOG's Contingent Bayes Networks. This allows greater flexibility in picking an ordering of variables to be sampled, and one planned use case of this functionality would be to select sampling orders that determine whether a partial world is unsupported sooner rather than later. The outcome of such an optimization algorithm would be to more rapidly discard partial worlds and thus improve the end-to-end performance of BLOG inference for a fixed desired number of samples.
- ii. Maintenance, testing and debugging of BLOG examples given recent syntax and language updates (e.g., array-valued variables, map data structures, etc.).
- iii. Profiling and experimentation of the revised BLOG examples under Java. The findings here reaffirmed the performance gap between the Java and C implementations, particularly in terms of data structure traversals and interpretation overheads in the Java version. Our findings motivated the following action item.
- iv. A code generator API design from BLOG models to allow the generation of both low-level C++, and for the development of a distributed multi-node BLOG runtime on K3. The API operates on model structures, defining an interface to translate dependency and number statements, as well as evidence statements and queries. Experiments with hand-compiled inference code for seven of the BLOG library models showed speedups of 30x to 100x compared to the original generic algorithm, while separate experiments demonstrated speeds of up to 60 million MCMC samples per second on a single core with carefully tuned model-specific C code. Based on this work, Berkeley successfully proposed development of compiler technology for probabilistic programming to the DARPA PPAML program.

After discussion across the JHU and UCB teams, we believe a code generation approach out of a common Java front end (that provides a parser and internal representation) will allow us to realize a scalable, yet integrated software implementation and architecture. In particular, we can generate both an efficient single machine inference engine that can exploit multicore execution with C++ and OpenMP, as well as a multi-machine implementation that can maximally reuse existing software for the networking, communication, and scheduling components.

As noted under Task 1, the BLOG language, inference engine, and inference server are available at <http://bayesianlogic.cs.berkeley.edu>

7 References

1. Lei Li and Stuart J. Russell, "*The BLOG Language Reference*". EECS Department, University of California, Berkeley, Technical Report UCB/EECS-2013-51, May, 2013.
2. Lei Li, Bharath Ramsundar and Stuart Russell, "*Dynamic Scaled Sampling for Deterministic Constraints*", In Sixteenth International Conference on Artificial Intelligence and Statistics, 2013.
3. Yusuf Erol, Lei Li, Bharath Ramsundar and Stuart J. Russell, "The Extended Parameter Filter", In Proceedings of the 30th International Conference on Machine learning, 2013.
4. Mark Rogers, Lei Li, Stuart Russell. Dynamic Tensor Systems for Weather Prediction. Workshop on Understanding Climate Change. 2012.
5. Sharad Vikram, Lei Li, and Stuart Russell. Handwriting and gestures in the air, recognizing on the fly. In CHI Extended Abstracts, 2013.
6. Shaunak Chatterjee and Stuart Russell, "Uncertain observation times." In *Proc. SUM-12*, Marburg, Germany, 2012.
7. Nick Hay, Stuart Russell, S. E. Shimony, and D. Tolpin, "Selecting Computations: Theory and Applications." In *Proc. UAI-12*, Catalina Island, 2012.
8. Siddharth Srivastava, Stuart Russell, Avi Pfeffer, "First-Order Models for POMDPs." In *Proc. 2nd International Workshop on Statistical Relational AI (StarAI-12)*, Catalina Island, 2012.
9. Edwin K.P. Chong and I-Jeng Wang, "On Policy Improvement of Rollout with Feature-Based Policies," in preparation for submission to *Operations Research Letters*, 2013.
10. N. Naikal, D. Singaraju, and S. Sastry, "Using models of objects with deformable parts for joint categorization and segmentation of objects", *Asian Conference on Computer Vision*, 2012.
11. H. Ohlsson, Y. Eldar, A. Yang, and S. Sastry, "Compressive shift retrieval", *ICASSP*, 2013.
12. L. Zhuang, A. Yang, Z. Zhou, S. Sastry, and Y. Ma, "Single-sample face recognition with image corruption and misalignment via sparse illumination transfer", *CVPR* 2013.
13. Allen Yang, Liansheng Zhuang, Zihan Zhou, Shankar Sastry, and Yi Ma, "Method and apparatus for single-image robust face alignment via illumination transfer", *US patent application*, 2013.
14. D. Rother, S. Schutz, R. Vidal. *Hypothesize and Bound: A Computational Focus of Attention Mechanism for Simultaneous n-d Segmentation, Pose Estimation and Classification Using Shape Priors*. arXiv: 1104.2580v2, 2011.
15. D. Rother and R. Vidal. A Hypothesize-and-Bound Algorithm for Simultaneous Object Classification, Pose Estimation and 3D Reconstruction from a Single 2D Image. In *ICCV Workshop on 3D Representation and Recognition*, 2011.
16. D. Rother, S. Mahendran, R. Vidal. *Hypothesize and Bound: A Computational Focus of Attention Mechanism for Simultaneous 3D Shape Reconstruction, Pose Estimation and*

- Classification from A Single 2D Image*. International Journal of Computer Vision, 2013. (under revision).
17. S. Mahendran, E. Yoruk, R. Vidal. *3D Object Classification, Pose Estimation and Reconstruction in Cluttered Images*. ICCV Workshop on 3D Reconstruction and Recognition, 2013 (under review).
 18. E. Yoruk, R. Vidal. *A Computationally Efficient Approach to 3D Model Matching for 3D Object Localization and Fine-Grained 3D Pose Estimation from a Single 2D Image*. ICCV Workshop on 3D Reconstruction and Recognition, 2013 (under review).
 19. D. Singaraju, and R. Vidal. *Global Bag-of-Features Models in Random Fields for Joint Categorization and Segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013 (under review)
 20. A. Jain, L. Zappella, P. McClure, and R. Vidal. Visual Dictionary Learning for Joint Object Categorization and Segmentation. European Conference on Computer Vision, 2012.
 21. A. Jain, S. Chatterjee, S. Russell, and R. Vidal. Coarse-to-fine Semantic Video Segmentation using Supervoxel Trees. International Conference on Computer Vision, 2013
 22. Bin Fu, Jiali Lin, Lei Li, Christos Faloutsos, Jason Hong and Norman Sadeh (2013), *"Why People Hate Your App - Making Sense of User Feedback in a Mobile App Store"*, In KDD '13: Proceeding of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA. ACM.
 23. Brian Rutenber and Avi Pfeffer, *"Probabilistic Decision Programming with Figaro"*, NIPS Workshop on Probabilistic Programming, 2012.

List of Acronyms, Abbreviations, and Symbols

Acronym	Description
BLOG	Bayesian Logic
BoF	Bag-of-Features
CRF	Conditional Random Field
DARPA	Defense Advanced Research Projects Agency
DBLOG	Temporal/Dynamic Extension of BLOG
DBN	Dynamic Bayesian Network
DYSC	Dynamic Scaling Algorithm
ECSS	Exact Constrained Sequential Sampling
EM	Expectation Maximization
H&B	Hypothesis-and-Bound Algorithm
HOG	Histogram of Oriented Gaussian
POMDP	Partially Observable Markov Decision Process
PP	Probabilistic Programming
SRC	Sparse-Representation Based Classification